CaseLog: Semantic Network Interface to a Student Computer-based Patient Record System

Christopher Cimino, M.D., Elizabeth K. Goldman, M.D., Jane A. Curtis, M.D., and Michael J. Reichgott, M.D., Ph. D., Office of Computer Based Education, Department of Pediatrics, and Office of the Dean of Students and Graduate Medical Education
Albert Einstein College of Medicine, Bronx, N.Y.

We have developed a computer program called CaseLog, which serves as an exemplary, computer-based patient record (CPR) system. The program allows for the introduction of the students to issues unique to patient record systems. These include record security, unique patient identifiers, and the use of controlled vocabularies. A particularly challenging aspect of the development of this program was allowing for student entry of controlled vocabulary terms. There were four goals we wished to achieve: students should be able to find the terms they are looking for; once a term has been found, it should be easy to find contextually related terms; it should be easy to determine that a sought-for term is not in the vocabulary; and the structure of the vocabulary should be dynamically altered by contextual information to allow its use for a variety of purposes. We chose a semantic network for our vocabulary structure. Within the processing power of the equipment we were working with, we achieved our goals. This paper will describe the development of the vocabulary, the design of the CaseLog program, and the feedback from student users of the program.

INTRODUCTION

Predictions about the Computer-Based Patient Record (CPR) have changed over the last several years. For many years it was predicted that a CPR would soon become a reality [1]. Now it is predicted that a CPR will soon become essential [2]. A challenge for medical educators is to familiarize students with the use of a CPR, since such programs may not now be generally available at clerkship sites.

There are several concepts related to a CPR that may not be evident, even to today's computer-literate students. These include patient record security, unique identification of individual patients, and controlled vocabularies. Of these, the use of controlled vocabularies is often the most difficult to understand.

Controlled Vocabularies

Controlled vocabularies are used to systematically index fluid English language information. A good controlled vocabulary will contain unambiguously defined terms without redundancy, will have an explicit set of relationships between terms, and will allow terms to appear in 'multiple contexts' in a consistent manner [3]. 'Multiple contexts' refers to the way in which a term is linked to other terms. For example, 'Viral Pneumonia' may be linked to the more general terms 'Pulmonary Diseases' and 'Infectious Diseases'. These are two of the 'contexts' in which 'Viral Pneumonia' exists. The collection of all the context links in the vocabulary define the structure of the vocabulary.

Design of a good controlled vocabulary may be at odds with good interface design. Redundant terms in the interface can be advantageous because they increase the chance that a user will find a needed term. A vocabulary designed for a user interface will try to balance the number of levels in the vocabulary with the number of concepts at each level so that a user avoids looking through long lists or making numerous selections to get to the term he or she wants. Sometimes it is useful to relate a very specific term with very general terms solely because the latter terms will be easy to find and the specific term is one that is frequently used. It may be useful to relate two terms because they frequently co-occur. These considerations are unrelated to explicit definitions or vocabulary structure based on consistent meaning. A semantic network structure can act as a bridge between a rigid controlled vocabulary and a more flexible interface vocabulary.

Semantic Networks

A semantic network is a knowledge representation scheme that defines concepts based on how they are linked to other concepts [4]. For example, an object may be defined by saying it is round, bouncy, and owned by Jack. In English we would say this is

0195-4210/92/$5.00

'Jack's ball'. Note that the other concepts (i.e., round, bouncy, ownership, and Jack) are also defined based on links to other objects in the network. For example, ownership might be linked to the concepts of buying and selling. Jack would be linked to the concept of people. Further links would show that owning, buying, and selling are things that people can do. Based on these links the semantic network can verify that Jack can own a ball.

There are several properties of semantic networks that are useful for our purposes. The first property is that even though there are several methods of describing a concept, the semantic network reduces these to a single internal representation. For example, 'Jack's ball' and 'the ball owned by Jack' are both valid but the system recognizes them as being the same object. In our vocabulary, this property allows seemingly redundant terms to be resolved into unambiguous terms. For example,'Oral Candidiasis' and 'Thrush' can both be internally represented as 'inflammation of the oral mucosa by Candida albicans'.

Another property of a semantic network is that objects can be closely linked to those objects that they most commonly co-occur with. For example, 'Immunizations' are considered 'Routine'. A 'Well Baby Exam' is also 'Routine'. As a result, these two concepts will be closely linked in the semantic network. Once the student finds one of these terms in the interface, the other term will be close by.

It is possible to filter the view of the semantic network. For example, all terms that are linked by the 'caused by' relation could be collected. This would produce a list of general terms like 'Allergy' or 'Infection' linked to terms like terms like 'Pollen' and 'Dust' or 'Pneumococcus' and 'Varicella zoster'. Another common link in semantic networks is the 'is a' link. For example, 'Pneumonia' 'is a' 'Pulmonary Disease'. By filtering the semantic network view with one or more link types, the CaseLog program presents the appearance of multiple vocabulary structures. For example, the user interface filters for both 'is a' and 'caused by' links. This allows many more co-occurring terms to be brought in than would otherwise exist in a standard hierarchy. A different filter set is used to produce student summary reports. The interface filter encourages redundancy in the vocabulary, making it easier to find desired terms. The summary report filter eliminates redundancy making reports more consistent and making comparison feasible for reports of data from different clerkship sites.

PROGRAM DESIGN

The flexibility of the semantic network structure allowed the program design to be based largely on faculty input. Initially, the terms for the vocabulary were created ad hoc by the clerkship directors. While they could then be edited to create a controlled vocabulary, the semantic network made it possible to reproduce the vocabulary in a form matching the directors' original submissions. It also allowed the vocabulary to be adapted to the interface and report formats.

None of the clerkship sites currently uses a CPR. Therefore, the next step was to create a mock-up so that clerkship directors would have some idea of how the system would operate and what its limitations would be. Comments about the mock-up elicited from the Pediatrics and Medicine clerkships played a role in the choice of software used for CaseLog. Specifically, a standard relational database package did not provide enough flexibility for term selection. Another important aspect to the mock-up was that it provided sample patient and student records and summary reports. This allowed the directors to see the role played by the controlled vocabulary terms they chose, from a student perspective as well as a faculty perspective.

Because of the hardware available, voice input, pen input, and other new technologies were not available for use in selecting vocabulary terms in CaseLog. This left two main alternatives; the student could type a term into the keyboard and review a list of possible matching terms or the student could browse through lists of terms from the controlled vocabulary. We chose to implement the latter alternative first. Such a browser usually implies that the vocabulary will be structured in a hierarchical fashion.

Platform

The initial CaseLog hardware platform was an Apple™ Macintosh SE. The programming environment chosen was Digitalk™ Smalltalk/V.

Interface

Record Security: A student starting CaseLog must first supply an identification number and password. Under some circumstances he will be asked to change his password and can't thereafter reuse the old password.

Unique Patient Identification: Once she has satisfied the security module, the student is presented with a window which displays her student record

(Figure 1). This includes a list of the patients she has seen. She can add a new patient or can use the mouse to edit the record of a previously seen patient.

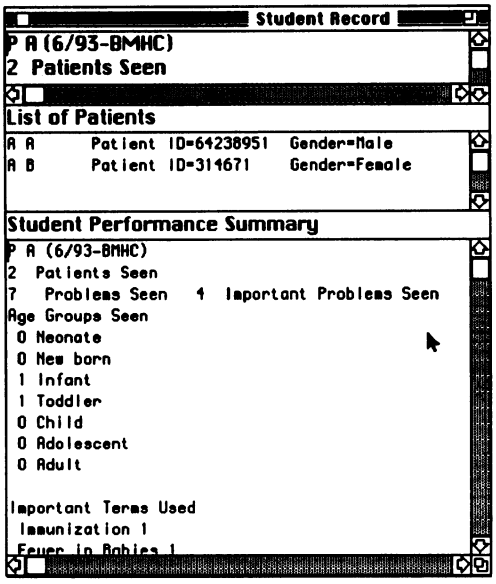

Figure 1: Student Record window showing student P.A.'s patient list and performance summary. (BMHC = Bronx Municipal Hospital Center)

When a student enters a new patient, he must supply a unique identification (ID) number (usually the hospital or clinic ID number), the patient's name, date of birth, and gender. If the unique ID number matches that of a patient the student has already seen, he is instructed to find the old record in his list of patients. Students can't see what patients other students have entered but if a student enters an ID that matches the ID of another patient in the system, the student is stopped if the name, date of birth, or gender do not match. The system also compares patient names. If a conflict is found (i.e., matching names but mismatched ID's), the student is warned but can proceed with creating a new record.

The patient record window displays the patient's identifying information (e.g., ID, name, gender), a chronic problem list, a list of visits, and information associated with each visit (Figure 2). Each visit has a date, a note, and an acute problem and visit term list. The visit note holds uncontrolled text. The student can change the visit date or the visit note directly on the patient record window. Visit terms include history, physical, and procedure terms. Chronic problems, acute problems, and visit terms are added to the lists with a vocabulary browser.

Controlled Vocabulary Selection: The size of the CaseLog vocabulary has changed throughout the program's use but it has approximately 600 terms. The majority of the terms are diagnoses or terms used to categorize other terms. The vocabulary browser is divided into three columns (Figure 3). Initially, the extreme right column contains a list of general categories. The focus of the window is the top center 'selection' box. When the student picks a term, it is moved into that box. The column on the left contains terms that are more general than the selected term. The column on the right contains terms that are more specific. The column below the selection box contains related terms and relies heavily on the semantic filtering of the vocabulary.

When the student moves the term he wants into the selection box, he clicks on the Select button and the term is transferred to the patient record. The browser remains open. The student may select as many terms as he wants before exiting from the browser. Once he finds the first term, many of the additional terms he wants will already be visible in the window.

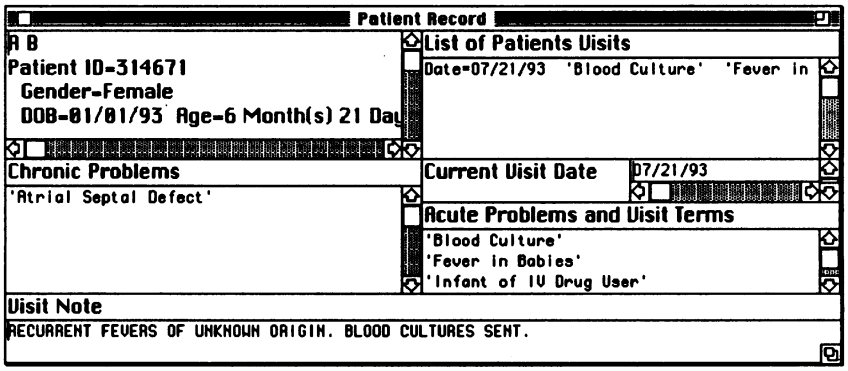

Figure 2: Patient Record window showing information for patient A.B.

METHODS

The CaseLog program was installed at an ambulatory pediatrics clerkship site in September of 1992. The program has been used by four rotations of 7 to 8 students per rotation at this site (total 30 students). Students were given a one hour orientation session at the beginning of each rotation with each student using the program for 10 to 15 minutes. Students reported prior experience with computers ranging from none to extensive programming experience. Three quarters of the students had used a word processor in the past. Slightly more than one third had prior experience using a Macintosh or a mouse. Students were asked to keep written records of their patients as well as enter the cases into CaseLog. At the end of each rotation, students were required to turn in their written records. All students were asked to hand in written comments about the CaseLog program. A non-random sample of students was interviewed concerning their use of the CaseLog program.

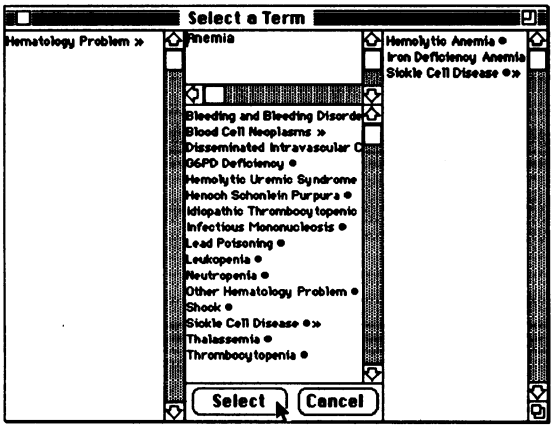

Figure 3: Vocabulary Selection window showing Anemia as the currently selected term. The left hand column shows more general terms, the right hand column more specific terms, and the middle column related terms. '»' indicates a term is linked to more specific terms. '•' indicates a term is specific enough to be used in a patient record.

Most students at the ambulatory site saw between 60 and 80 cases. Students entered into CaseLog an average of 47 cases or 61% of the patients they had actually seen. One student entered no cases and could not be reached for comments. All other students entered at least 15 cases and 6 students entered all their cases.

Case Entry

One hundred and sixteen cases were reviewed, four for each student, and compared to the written records. Students entered an average of 0.3 chronic problems per patient (range 0-2) and 5.6 visit terms (range 0-13). Almost all of the information that appeared in the written notes was duplicated in CaseLog either as a controlled term or in the uncontrolled visit note. The majority of material in the computer visit notes described drugs prescribed, referrals made, and follow-up plans. A total of 37 terms were found (0-2 per case) in the visit notes that could have been added as controlled terms. Most of these dealt with patient or parent education.

User Feedback

The chief criticism of the program was that it was too slow. The system slowed down appreciably as the number of cases entered increased.

The second most common criticism was that students could not find the appropriate terms. Twenty-seven such "missing" terms were reported. Three of the 27 terms cited were in the vocabulary but not where the student was looking for them. The other 24 terms were missing from the vocabulary. These 27 complaints resulted in 22 terms being added and 2 terms being relocated. The 8 students at the ambulatory site in the most recent rotation did not report any missing terms.

The third most common complaint was that students wanted the addition of a 'look-up' feature. None of the students complained that the program was difficult to use. Eighteen students who were questioned directly, including 12 who described themselves as novice users, all stated the program was easy to use and they had no difficulty with it.

DISCUSSION

The results described are mostly subjective but seem to show that CaseLog was easy to use, that it was used by most of the students to enter most of their cases, that students successfully located appropriate terms, and successfully captured most of the information contained in their cases. The latter two interpretations must be made with caution because although the spot comparison of written records to CaseLog records seems to show completeness, it is possible students chose terms for their written record based on familiarity with the CaseLog vocabulary. In other words, after using CaseLog and becoming familiar with its vocabulary, they may have started using the same vocabulary for their written records. It is also likely that students settled for less than perfect matches since this would be easier than either complaining about the missing term or entering the term in the uncontrolled portion of the visit note.

A second benefit of students using a CPR is that the data collected can provide valuable information about the types of medical problems and procedures to which the students are exposed [5, 6]. While there is no guarantee that exposure leads to learning, it is possible to detect important concepts about which the students are not learning. In the course of a clerkship, it is assumed that students will see a broad selection of important problems pertaining to a particular specialty but there is no way to control that. If the clerkship directors knew that particular students had not seen a particular type of problem, they could supplement the group's learning with a didactic session or case simulation.

The Albert Einstein College of Medicine recently reorganized its third year Pediatrics Clerkship [7]. One of the results of this reorganization was that some of the clerkship sites would continue to have a traditional in-patient focus and other sites would implement an innovative out-patient focus. A third benefit of a CPR in this situation will be to collect data so these two types of clerkship experiences can be compared.

It should be noted that CaseLog presents the student with a program that has the 'look and feel' of a CPR. As such, it is not a substitute for the traditional history and physical exam write-up. In particular, the structure of the CaseLog vocabulary encourages bad habits such as using 'Normal Exam'. On the other hand, CaseLog discourages the use of abbreviations because all its terms are explicitly described.

Conclusion

CaseLog has provided two of the planned benefits: students are exposed to a CPR and basic data is collected about the patients students see. After almost a year of use at the ambulatory clerkship site, CaseLog has recently been installed at one of the in-patient clerkship sites. Students at this site have already begun identifying terms that are missing from the vocabulary because they are not common to an ambulatory setting (i..e., 'Wilm's Tumor'). It is anticipated that it will require 2 or 3 in-patient rotations before the vocabulary again becomes stable. It is also anticipated that as student records from these different sites are compared, new filters will be developed to highlight the differences and similarities. It will be interesting to see if the semantic network structure continues to be useful as the CaseLog vocabulary grows beyond its present size.

Reference

[1]. McDonald, Clement J., Tierney, William M. Computer-Stored Medical Records: Their Future Role in Medical Practice. JAMA 1988; 259(23):3433-3440.

[2]. The Computer-based Patient Record: An Essential Technology for Health Care. [Dick, Richards S., Steen, Elaine B., eds.] The Institute of Medicine, National Academy Press. 1991.

[3]. Cimino, James J., Hripcsak, George, Johnson, Stephen B., Clayton, Paul D. Designing an Introspective, Multipurpose Controlled Medical Vocabulary. In: Symposium on Computer Application in Medical Care [Kingsland, Lawrence C. ed]. IEEE Computer Society Press, Washington, 1989; 513-518.

[4]. Woods, William A. Understanding Subsumption and Taxonomy: A Framework for Progress. In: Principles of Semantic Networks: Explorations in the Representation of Knowledge [Sowa, John, ed.]. Morgan Kaufman, 1991.

[5]. Oliver, Diane E., Piggins, Judy, Hassan Laurie, Ryan Blewett, Dyan, Eccles, Randy, Schoelles, Craig, Raila, Wayne, Link, David, Barnett, G. Octo. Casebook: A system for tracking cases seen by medical students. In: Proceedings of 1st American Medical Informatics Educational and Research Conference [Mitchell, Joyce A. ed.]. Apple Computer, Snowbird Utah, 1990: page 10.

[6]. Kahn, Joel A., Piggins, Judith, Ryan Blewett, Dyan Ryan, Hassan, Laurie, Raila, Wayne, Link, David, Oliver, Diane E., Barnett, G. Octo. Casebook: A System for Tracking Clinical Encounters. In: Symposium on Computer Applications in Medical Care [Clayton, Paul D., ed.] McGraw Hill, New York, 1991: 718-722.

[7]. Goldman, Elizabeth K., Reichgott, Michael J., Stein, R.E.K. A pediatric clerkship based in practice teams. In: Conference on Research in Medical Education, American Association of Medical Colleges, New Orleans, 1992.